

Graph Regularized Low Rank Representation for Aerosol Optical Depth Retrieval

Yubao Sun, Renlong Hang, Qingshan Liu, Fuping Zhu, Hucheng Pei

Abstract

In this paper, we propose a novel data-driven regression model for aerosol optical depth (AOD) retrieval. First, we adopt a low rank representation (LRR) model to learn a powerful representation of the spectral response. Then, graph regularization is incorporated into the LRR model to capture the local structure information and the nonlinear property of the remote-sensing data. Since it is easy to acquire the rich satellite-retrieval results, we use them as a baseline to construct the graph. Finally, the learned feature representation is feeded into support vector machine (SVM) to retrieve AOD. Experiments are conducted on two widely used data sets acquired by different sensors, and the experimental results show that the proposed method can achieve superior performance compared to the physical models and other state-of-the-art empirical models.

Keywords

Aerosol optical depth (AOD), physical model, empirical model, low rank representation (LRR), graph regularization, support vector machine (SVM).

I. INTRODUCTION

Aerosols are small solid and liquid particles suspended in the atmosphere. They can scatter and absorb solar radiance, and modify microphysical and radiative properties of cloud [1]. An important metric of aerosols' concentration in the atmosphere is aerosol optical depth (AOD), which measures the amount of depletion that a beam of solar radiation undergoes as it passes through the atmosphere. AOD has become a major atmospheric data product derived from various earth observation satellites, such as the Moderate Resolution Imaging Spectroradiometer (MODIS) [2], the Multiangle Imaging Spectro-Radiometer (MISR)

Y. Sun, R. Hang and Q. Liu are with the Jiangsu Key Laboratory of Big Data Analysis Technology, the School of Information and Control, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: sunyb@nuist.edu.cn; renlong_hang@163.com; qslu@nuist.edu.cn).

F. Zhu and H. Pei are with Beijing Electro-Mechanical Engineering Institute, Beijing 100083, China.
Corresponding author: Qingshan Liu

[3], etc. These remote sensing satellites obtain geolocated and calibrated radiances, and then retrieval algorithms are used to derive the corresponding AOD values.

Most of the current operating satellite-retrieval algorithms are based on physical models [4][5][6]. These models need to take into account numerous physical variables affecting the radiometric characteristics of remote sensing data, such as atmospheric conditions, solar azimuth and zenith angles, sensor azimuth and zenith angles, etc. Complex mathematical formulations are set up to represent the relationships between these variables according to radiation transfer equation. To simplify the radiative transfer calculations, a lookup table (LUT) is used to simulate the radiative properties of the atmosphere calculated for expected aerosol types at particular wavelengths, angles and aerosol loading. Spectral reflectance from the LUT is then compared with the satellite-observation value to find the best match, and the corresponding AOT is the final retrieval result. However, due to the complex Earth-atmosphere interaction, it is difficult to consider all related physical variables and to accurately formulate their relationships. Besides, searching LUT is very time consuming.

An efficient alternative is empirical models [7][8]. These models can be considered as data-driven regression approaches. First, using the collocated satellite and ground-based observations to train a regression model. Then, the trained model is used to predict AOD for satellite observations without ground-truth. Among these models, neural networks (NN) [8][9] and support vector machines (SVMs) [7][10] are two popular methods, because they can approximate the complex non-linear relationships between satellite-based observations and ground-based observations. In contrast to physical models, empirical models don't make *a priori* assumptions on variable relations or rigid functional forms, they directly rely on the available data. In addition, they are computationally less expensive, flexible to different retrieval scenarios and more accurate than the physical models, if sufficient amounts of training data are available. However, in previous works, most empirical models directly exploit a part of or the whole spectral values as features. Since remote-sensing data often suffer from various annoying degradations, e.g., noise contamination, stripe corruption, and missing data, due to the sensor, photon effects, and calibration error [11][12], directly using such

corrupted data without any preprocessing may degrade the performance of empirical models.

In this paper, we propose a graph regularized low rank representation model to address the above issues. Low rank representation (LRR), which was successfully employed in natural image denoising [13][14], has shown its potential in remote-sensing data analysis [15]. By stacking the remote-sensing data into a 2-D matrix, it should be low rank due to the high correlations of the spectral information. Moreover, as discussed in [16], the rank of the matrix constructed by remote-sensing data is bounded by the small number of pure spectral endmembers. We therefore propose to employ LRR to learn a new feature representation for the remote-sensing data. Besides, due to the effects of multipath scattering, variations in sun-canopy-sensor geometry, nonhomogeneous composition of pixels, and attenuating properties of media, remote-sensing data are often nonlinearly distributed [17]. To preserve such nonlinear structure, motivated by [18][19][20], graph regularization is incorporated into the LRR model. Nowadays, it is easy to acquire the rich satellite-retrieval results based on physical models. Many researchers attempted to employ such information as features for regression models [7][9][21]. Different from them, we use the satellite-retrieval results as a baseline to construct the graph. This prior information can make the proposed model combine the merits of physical models and regression models to some extent.

The rest of this paper is outlined as follows. Section II introduces the proposed method in detail. Section III presents the data sets used and the experimental results, followed by the conclusion in Section IV.

II. METHODOLOGY

A. Representation model

Assume the remote-sensing data can be represented as a two-dimensional matrix $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n] \in \mathbf{R}^{d \times n}$ by stacking the pixels in the original d -dimensional spectral feature space as the columns. Since the remote-sensing data are often corrupted by noise in the acquisition process, \mathbf{X} can be written as $\mathbf{X} = \mathbf{X}_0 + \mathbf{E}$, where $\mathbf{X}_0 \in \mathbf{R}^{d \times n}$ is the ideal clean data, $\mathbf{E} \in \mathbf{R}^{d \times n}$ is the noise or outlier. Each pixel of \mathbf{X}_0 can be represented by a linear combination of the bases in a ‘dictionary’ $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m] \in \mathbf{R}^{d \times m}$.

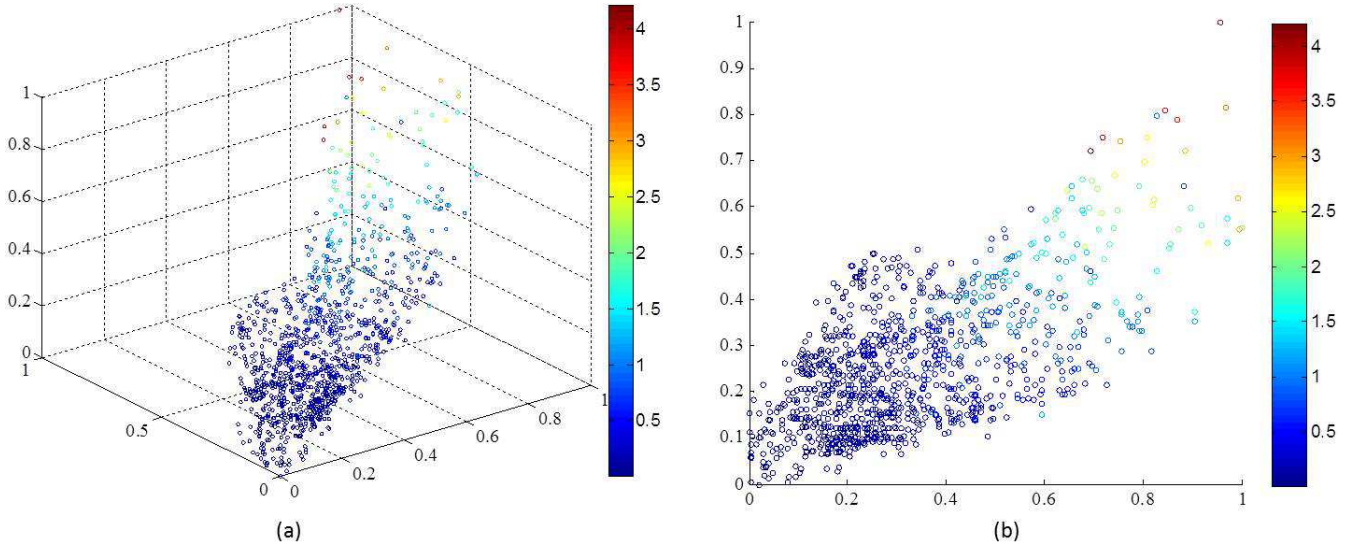


Fig. 1: An example of nonlinear distribution of MISR remote-sensing data. (a) A three-dimensional map by choosing three spectral bands in MISR. (b) A two-dimensional map of (a).

So we have $\mathbf{X} = \mathbf{AZ} + \mathbf{E}$, where $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_i, \dots, \mathbf{z}_n] \in \mathbf{R}^{m \times n}$ is the coefficient matrix and each \mathbf{z}_i corresponds to the new characterization of \mathbf{x}_i . The dictionary is often overcomplete, and the clean remote-sensing data often lie in a low rank feature space [16]. Thus, we can use an alternative scheme of $\mathbf{A} = \mathbf{X}$ as in [13][22], and the purpose becomes to search for the lowest rank solution of \mathbf{Z} by

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{E}} \quad & \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{2,1}, \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{XZ} + \mathbf{E}, \end{aligned} \tag{1}$$

where $\|\cdot\|_*$ denotes the *nuclear norm* of a matrix, and \mathbf{E} represents the sparse noise, which is measured by $l_{2,1}$ norm, i.e., $\|\mathbf{E}\|_{2,1} = \sum_{j=1}^n \sqrt{\sum_{i=1}^d (\mathbf{E}_{ij})^2}$ as in [14]. $\lambda > 0$ is a balance parameter. This is a popular LRR model. It was demonstrated that LRR is capable of capturing the global structure of the data as well as robust to noise [22].

However, as discussed in [23], nonlinearities often exist in remote-sensing data due to the effects of many factors. Taking MISR remote-sensing data as an example in Fig. 1. Fig. 1(a) shows a three-dimensional map by choosing three spectral bands in MISR. Obviously, this data is nonlinear distribution, which can also be demonstrated from its two-dimensional map. Recently, nonlinear manifold learning has been proved

to be a successful method to capture such nonlinearity [24][25]. Inspired by the idea, we propose a graph regularization based LRR model to preserve the local neighboring relations among the data. Thus, the objective function in Eq. (1) can be reformulated as

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{E}} \quad & \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{2,1} + \frac{\beta}{2} \text{tr}(\mathbf{Z}\mathbf{L}\mathbf{Z}^\top), \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}. \end{aligned} \quad (2)$$

The term of $\text{Tr}(\mathbf{Z}\mathbf{L}\mathbf{Z}^\top)$ is the graph regularization term, which is derived as follows:

$$\begin{aligned} \min_{\mathbf{Z}} \quad & \sum_{i=1}^n \sum_{j=1}^n \|\mathbf{z}_i - \mathbf{z}_j\|^2 w_{ij} \\ &= \min_{\mathbf{Z}} \text{Tr}(\mathbf{Z}(\mathbf{D} - \mathbf{W})\mathbf{Z}^\top), \\ &= \min_{\mathbf{Z}} \text{Tr}(\mathbf{Z}\mathbf{L}\mathbf{Z}^\top), \end{aligned} \quad (3)$$

where \mathbf{W} is an affinity matrix, \mathbf{D} is a diagonal matrix whose elements equal to the sum of rows or columns of \mathbf{W} , i.e., $\mathbf{D}_{ii} = \sum_{j=1}^n w_{ij}$, and $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the graph Laplacian matrix.

An intuitive motivation behind Eq. (3) is that if two pixels are close, their new representations are also close to each other [26]. As we all know, it is easy to acquire the rich satellite-retrieval results, we therefore use them as a baseline to construct the graph. Assume that the satellite-retrieval results of two pixels \mathbf{x}_i and \mathbf{x}_j are y_i and y_j , respectively, each element w_{ij} of \mathbf{W} can be calculated as

$$w_{ij} = \begin{cases} e^{-\frac{(y_i - y_j)^2}{2\sigma^2}}, & \text{if } y_i \in N_k(y_j) \text{ or } y_j \in N_k(y_i), \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

where $N_k(y_j)$ denotes the set of k -nearest neighbors of y_j , and σ refers to the parameter of the Heat kernel. This weight setting promotes our model to be consistent with the satellite-retrieval results, which can inherit the merits of both physical model and regression model.

B. Optimization algorithm

The objective function Eq. (2) of the graph regularized LRR is non-convex, thus jointly optimizing \mathbf{Z} and \mathbf{E} is extremely difficult. As in [27][28], we adopt the linearized Alternating Direction Method of Multipliers (ADMM) algorithm to optimize it. We first convert (2) to the following equivalent problem:

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{E}, \mathbf{J}} \quad & \|\mathbf{J}\|_* + \lambda \|\mathbf{E}\|_{2,1} + \frac{\beta}{2} \text{tr}(\mathbf{Z}\mathbf{L}\mathbf{Z}^\top), \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}, \quad \mathbf{Z} = \mathbf{J}, \end{aligned} \quad (5)$$

which can be changed to the Augmented Lagrange Multiplier (ALM) problem:

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{E}, \mathbf{J}} \quad & \|\mathbf{J}\|_* + \lambda \|\mathbf{E}\|_{2,1} + \frac{\beta}{2} \text{tr}(\mathbf{Z}\mathbf{L}\mathbf{Z}^\top) + \langle \mathbf{Y}_1, \mathbf{Z} - \mathbf{J} \rangle + \\ & \langle \mathbf{Y}_2, \mathbf{X}\mathbf{Z} + \mathbf{E} - \mathbf{X} \rangle + \frac{\mu}{2} \|\mathbf{X}\mathbf{Z} + \mathbf{E} - \mathbf{X}\|_F^2 + \frac{\mu}{2} \|\mathbf{Z} - \mathbf{J}\|_F^2, \end{aligned} \quad (6)$$

where $\mathbf{Y}_1, \mathbf{Y}_2$ are Lagrange multipliers, $\mu > 0$ is a penalty parameter, and $\|\cdot\|_F$ denotes the Frobenius norm. Eq. (6) can be rewritten as

$$\min_{\mathbf{Z}, \mathbf{E}, \mathbf{J}} \quad \|\mathbf{J}\|_* + \lambda \|\mathbf{E}\|_{2,1} + \frac{\beta}{2} \text{tr}(\mathbf{Z}\mathbf{L}\mathbf{Z}^\top) + \frac{\mu}{2} \|\mathbf{X}\mathbf{Z} + \mathbf{E} - \mathbf{X} + \frac{\mathbf{Y}_2}{\mu}\|_F^2 + \frac{\mu}{2} \|\mathbf{Z} - \mathbf{J} + \frac{\mathbf{Y}_1}{\mu}\|_F^2. \quad (7)$$

Following the iterative optimization method in [13], we divide Eq. (7) into three sub-problems: optimizing \mathbf{J} while fixing \mathbf{Z} and \mathbf{E} , optimizing \mathbf{E} while fixing \mathbf{J} and \mathbf{Z} , and optimizing \mathbf{Z} while fixing \mathbf{J} and \mathbf{E} .

Fixing \mathbf{Z} and \mathbf{E} to optimize \mathbf{J} , Eq. (7) is simplified to

$$\min_{\mathbf{J}} \quad \|\mathbf{J}\|_* + \frac{\mu}{2} \|\mathbf{J} - (\mathbf{Z} + \frac{\mathbf{Y}_1}{\mu})\|_F^2. \quad (8)$$

According to [29] [30], singular value thresholding operator is used to solve Eq. (8).

Fixing \mathbf{J} and \mathbf{Z} to optimize \mathbf{E} , Eq. (7) is reduced to

$$\min_{\mathbf{E}} \quad \lambda \|\mathbf{E}\|_{2,1} + \frac{\mu}{2} \|\mathbf{E} - (\mathbf{X} - \mathbf{X}\mathbf{Z} - \frac{\mathbf{Y}_2}{\mu})\|_F^2. \quad (9)$$

According to the **Lemma 3.2** in [13], if the optimal solution is \mathbf{E}^* , the i -th column of \mathbf{E}^* is:

$$\mathbf{E}^*(:, i) = \begin{cases} \frac{\|\mathbf{q}_i\| - \frac{\lambda}{\mu}}{\|\mathbf{q}_i\|} \mathbf{q}_i, & \text{if } \frac{\lambda}{\mu} < \|\mathbf{q}_i\|, \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

where \mathbf{q}_i is the i -th column vector of matrix $\mathbf{X} - \mathbf{XZ} - \frac{\mathbf{Y}_2}{\mu}$.

Fixing \mathbf{J} and \mathbf{E} to optimize \mathbf{Z} , Eq. (7) is simplified to

$$\min_{\mathbf{Z}} \frac{\beta}{2} \text{tr}(\mathbf{ZLZ}^\top) + \frac{\mu}{2} \|\mathbf{XZ} + \mathbf{E} - \mathbf{X} + \frac{\mathbf{Y}_2}{\mu}\|_F^2 + \frac{\mu}{2} \|\mathbf{Z} - \mathbf{J} + \frac{\mathbf{Y}_1}{\mu}\|_F^2. \quad (11)$$

We adopt a linearization strategy to optimize Eq. (11). In specific, $\frac{\mu}{2} \|\mathbf{XZ} + \mathbf{E} - \mathbf{X} + \frac{\mathbf{Y}_2}{\mu}\|_F^2$ is linearly approximated into the following formula by using second-order Taylor expansion around point \mathbf{Z}_k :

$$\frac{\mu}{2} \|\mathbf{XZ} + \mathbf{E} - \mathbf{X} + \frac{\mathbf{Y}_2}{\mu}\|_F^2 \approx \langle \mu \mathbf{A}^\top (\mathbf{XZ}_k + \mathbf{E} - \mathbf{X} + \frac{\mathbf{Y}_2}{\mu}), \mathbf{Z} - \mathbf{Z}_k \rangle + \frac{\mu \|\mathbf{A}\|_F^2}{2} \|\mathbf{Z} - \mathbf{Z}_k\|_F^2. \quad (12)$$

Then, substitute (12) for $\frac{\mu}{2} \|\mathbf{XZ} + \mathbf{E} - \mathbf{X} + \frac{\mathbf{Y}_2}{\mu}\|_F^2$ and Eq. (11) is written as follows:

$$\min_{\mathbf{Z}} \frac{\beta}{2} \text{tr}(\mathbf{ZLZ}^\top) + \frac{\mu \|\mathbf{X}\|_F^2}{2} \|\mathbf{Z} - (\mathbf{Z}_k - \frac{\mathbf{X}^\top (\mathbf{XZ}_k + \mathbf{E} - \mathbf{X} + \frac{\mathbf{Y}_2}{\mu})}{\|\mathbf{X}\|_F^2})\|_F^2 + \frac{\mu}{2} \|\mathbf{Z} - \mathbf{J} + \frac{\mathbf{Y}_1}{\mu}\|_F^2. \quad (13)$$

Finally, we can achieve the optimal \mathbf{Z} by setting the derivative of Eq. (13) with respect to \mathbf{Z} to zero:

$$\mathbf{Z} = [\mu \|\mathbf{X}\|_F^2 \mathbf{Z}_k - \mu \mathbf{X}^\top \mathbf{XZ}_k - \mu \mathbf{X}^\top \mathbf{E} + \mu \mathbf{X}^\top \mathbf{X} - \mathbf{X}^\top \mathbf{Y}_2 + \mu \mathbf{J} - \mathbf{Y}_1] (\beta \mathbf{L} + \mu (\|\mathbf{X}\|_F^2 + 1) \mathbf{I})^{-1}. \quad (14)$$

The detailed optimization algorithm is summarized in **Algorithm 1**.

Algorithm 1: The proposed optimization algorithm for graph regularized LRR by linearized ADMM

Input: Data matrix \mathbf{X} , parameter λ and β .

Initialize: $\mathbf{Z} = \mathbf{J} = \mathbf{0}$, $\mathbf{E} = \mathbf{0}$, $\mathbf{Y}_1 = \mathbf{0}$, $\mathbf{Y}_2 = \mathbf{0}$, $\mu = 10^{-6}$, $\max_\mu = 10^{11}$, $\rho = 1.1$, $\varepsilon = 10^{-11}$, $\mathbf{Z}_k = \mathbf{0}$.

Output: Lowest rank representation \mathbf{Z} .

```

1 for each iteration do
2   Fix  $\mathbf{Z}$ ,  $\mathbf{E}$  and update  $\mathbf{J}$ :  $\mathbf{J} = \operatorname{argmin} \|\mathbf{J}\|_* + \frac{\mu}{2} \|\mathbf{J} - (\mathbf{Z} + \frac{\mathbf{Y}_1}{\mu})\|_F^2$ .
3   Fix  $\mathbf{J}$ ,  $\mathbf{Z}$  and update  $\mathbf{E}$ :  $\mathbf{E} = \operatorname{argmin} \lambda \|\mathbf{E}\|_{2,1} + \frac{\mu}{2} \|\mathbf{E} - (\mathbf{X} - \mathbf{X}\mathbf{Z} - \frac{\mathbf{Y}_2}{\mu})\|_F^2$ .
4   Fix  $\mathbf{J}$ ,  $\mathbf{E}$  and update  $\mathbf{Z}$ :  $\mathbf{Z} =$ 
       $[\mu \|\mathbf{X}\|_F^2 \mathbf{Z}_k - \mu \mathbf{X}^\top \mathbf{X} \mathbf{Z}_k - \mu \mathbf{X}^\top \mathbf{E} + \mu \mathbf{X}^\top \mathbf{X} - \mathbf{X}^\top \mathbf{Y}_2 + \mu \mathbf{J} - \mathbf{Y}_1](\beta \mathbf{L} + \mu(\|\mathbf{X}\|_F^2 + 1)\mathbf{I})^{-1}$ .
5   Update the multipliers:  $\mathbf{Y}_1 = \mathbf{Y}_1 + \mu(\mathbf{Z} - \mathbf{J})$ ;  $\mathbf{Y}_2 = \mathbf{Y}_2 + \mu(\mathbf{X}\mathbf{Z} + \mathbf{E} - \mathbf{X})$ .
6   Update the parameters:  $\mu = \min(\rho\mu, \max_\mu)$ ;  $\mathbf{Z}_k = \mathbf{Z}$ .
7   Check the convergence conditions:  $\|\mathbf{X} - \mathbf{X}\mathbf{Z} - \mathbf{E}\|_\infty < \varepsilon$  and  $\|\mathbf{Z} - \mathbf{J}\|_\infty < \varepsilon$ .

```

III. EXPERIMENTS

A. Data Sets

As in [8][9][10], we use Level 2.0 AERONET retrievals as the target values for regression models. AERONET is a global network of about 250 ground-based instruments that observe aerosols [31]. Most of these stations measure AOD in different spectral bands centered around the nominal wavelengths of 340, 380, 440, 670 *nm*, and others [32]. To facilitate inter-comparisons with other instruments, these data are interpolated to 550 *nm* using the quadratic fit on log-log scale from all wavelengths, at a particular location and time [2]. Besides using the AOT values at 550 *nm* as ground-truth of regression problems, the following two sensors' spectral values along all the bands are used as features (inputs).

The first is the Moderate Resolution Imaging Spectroradiometer (MODIS) data. MODIS is a key instrument aboard the TERRA satellite for the collection of aerosol and cloud information. It has a swath width of 2330 *km*, and achieves a global coverage in about two days. The MODIS instrument has a single camera observing the top-of-the-atmosphere reflectance over 36 spectral bands between 410 *nm* and 14 μm at three different spatial resolutions (250 *m*, 500 *m*, 1 *km*) [33]. We obtain the MODIS Level-1B calibrated radiance product MOD021KM with spatial resolution of 1 *km*, covering the Beijing AERONET location between January 2002 and December 2014. Over the same spatial and temporal range, we obtain the Level-2 aerosol-retrieval product MOD04 (Collection 6, QA>1) with a spatial resolution of 10 *km*,

and geolocation product MOD03 with 1 km resolution. MOD04 product is used as a baseline to verify the effectiveness of regression models. Thereafter, Level 2.0 AERONET data are collocated in space and time with the MODIS data. The detailed process can be found in [9]. We obtain a total of 843 spatially and temporally collocated observations from MODIS and AERONET.

The second is Multi-angle Imaging SpectroRadiometer (MISR) data. MISR is one of the five instruments mounted on Terra spacecraft. The spacecraft flies in a sun-synchronous 705 km descending polar orbit, so that it crosses the equator always at 10:30 *am* local time. The MISR instrument consists of nine pushbroom cameras arranged in different view angles relative to the earth's surface. Each camera uses four Charge-Coupled Device (CCD) line arrays in a single focal plane. The line arrays cover 360 km wide swath and provide four spectral bands in Blue, Green, Red and Near Infrared (NIR) that are centered at 443, 555, 670 and 865 nm , respectively. The resolution of all the four bands in nadir view and the red band at all the nine angles is 275 m and the resolution of the other bands is 1.1 km . We download 1045 collocated MISR and AERONET data from MAPSS [32], covering the whole 23 stations at all available time in China.

B. Experimental Setup

To demonstrate the superiority of the proposed graph regularized LRR model, the new representation is feeded into the subsequent regression model SVM. For simplicity, we name it GLRR+SVM, which is compared with the following six retrieval models: 1) the operating satellite-retrieval algorithms by physical models; 2) the ordinary least square regression (OLS); 3) ridge regression (RR); 4) NN; 5) SVM; 6) the classical LRR with SVM regressor (LRR+SVM). For NN, the optimal number of hidden nodes is chosen from [2, 50] in steps of 5 via a 5-fold cross validation. For SVM, we adopt the Gaussian kernel since it usually achieves the best results compared to other kernels. The optimal variance parameter γ for the Gaussian kernel and the regularization parameter C in SVM are both selected from $\{10^{-3}, 10^{-2}, \dots, 10^3\}$. Besides, there are two regularization parameters λ and β for the graph regularized LRR model, which are also chosen by a 5-fold cross validation from the given set $\{10^{-3}, 10^{-2}, \dots, 10^3\}$. In all the experiments, we randomly divide the whole data into the training set and the testing set according to some percentages. The

training set is used to train the regression based retrieval models, while the testing set is used to evaluate their performances. In order to reduce the effects of random selection, all the algorithms are repeated ten times and the average performances are reported. Without loss of generality, we use two mainstream evaluation metrics: the root-mean-square error (RMSE) to evaluate the accuracy of the estimations, and Pearson's correlation coefficient R to evaluate the goodness of fit.

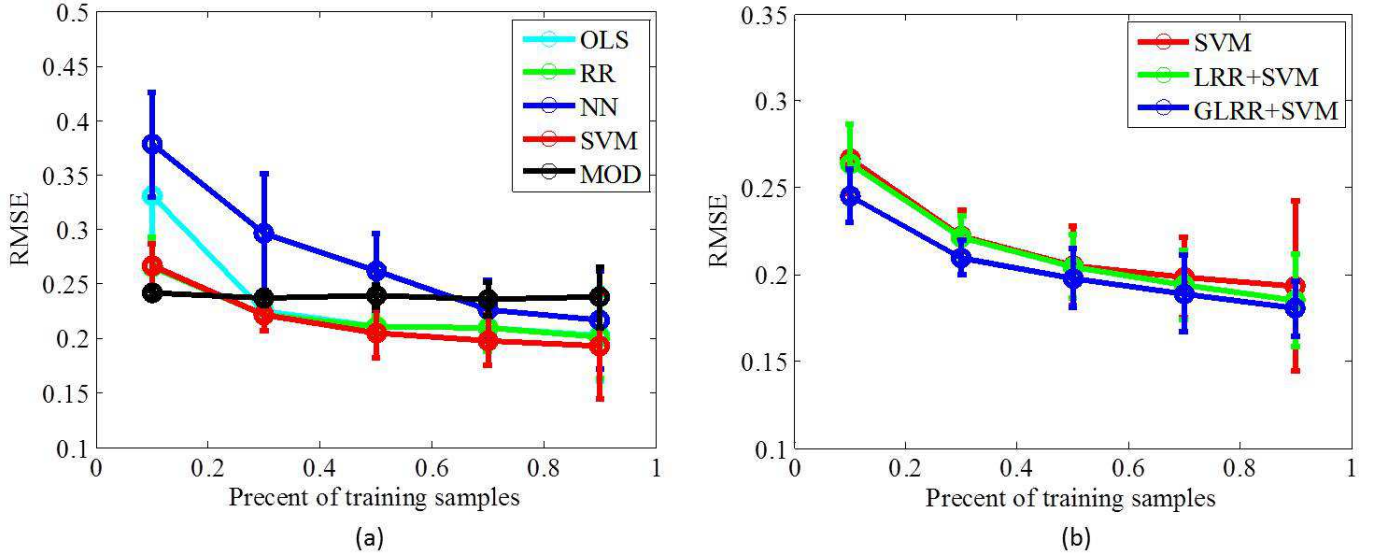


Fig. 2: RMSE and standard deviations achieved by seven different methods on MODIS data set. (a) Comparisons among SVM and four other methods. (b) Comparisons among SVM, LRR+SVM and GLRR+SVM.

C. Results and Discussion

For the MODIS data set, the average RMSE values and the standard deviations achieved by seven different models from ten experiments are demonstrated in Fig. 2, where the smaller values correspond to the better performances. Several conclusions can be observed from this figure. First, in Fig. 2(a), it can be seen that given enough training samples, the regression models achieve higher performance than the physical model. Second, SVM yields the best performance in the four regression models, because it can well solve the case with small numbers of training samples. Third, in Fig. 2(b), the performance of LRR+SVM is a little better than SVM especially when the percentage of training samples is more than 70%. This indicates that LRR is able to learn a better representation from the corrupted observation data compared to the pure

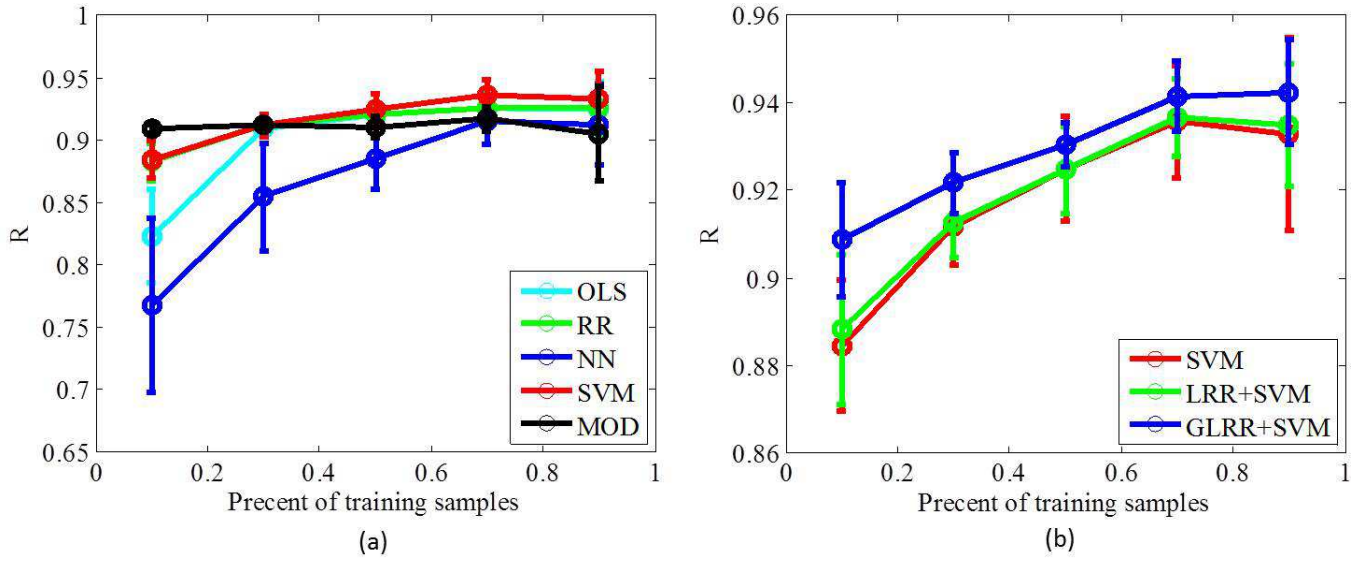


Fig. 3: R and standard deviations achieved by seven different methods on MODIS data set. (a) Comparisons among SVM and four other methods. (b) Comparisons among SVM, LRR+SVM and GLRR+SVM.

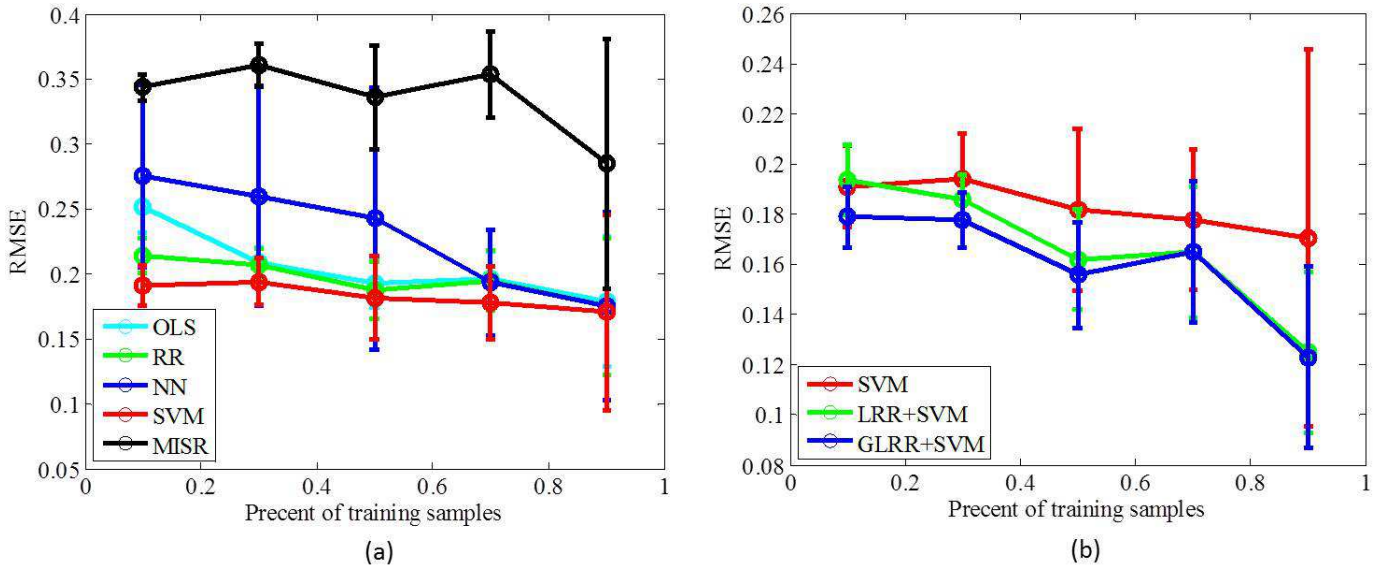


Fig. 4: RMSE and standard deviations achieved by seven different methods on MISR data set. (a) Comparisons among SVM and four other methods. (b) Comparisons among SVM, LRR+SVM and GLRR+SVM.

spectral response values. More importantly, GLRR+SVM further improves the performance as compared to LRR+SVM, because it can capture the local structure information and the nonlinear property of MODIS data. The last but not the least, a deficiency of regression models can be observed in Fig. 2(a). Specifically, when the percentage of training samples is 10%, the physical model is better than all of the regression

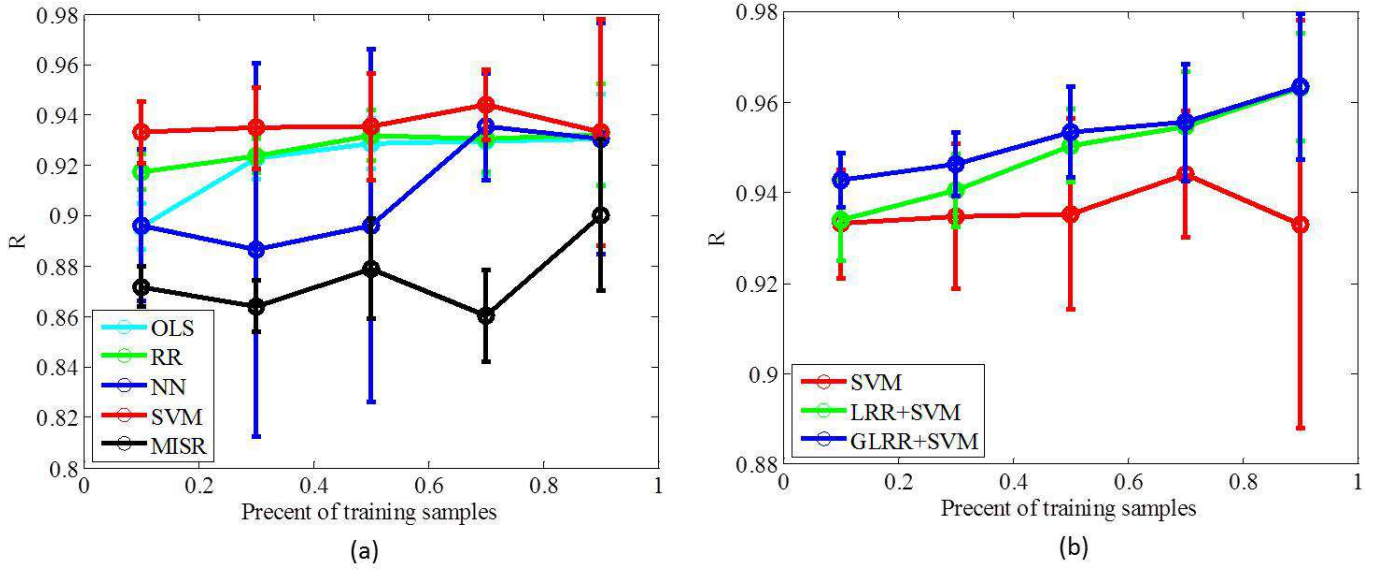


Fig. 5: R and standard deviations achieved by seven different methods on MISR data set. (a) Comparisons among SVM and four other methods. (b) Comparisons among SVM, LRR+SVM and GLRR+SVM.

models. However, when the percentage of training samples is more than 20%, the best regression model SVM yields higher performance than the physical model. This indicates that the regression models heavily rely on the number of training samples while the physical model is stable. The above conclusions can also be verified from another evaluation indicator R in Fig. 3. Different from Fig. 2, here, the larger values denote the better performances. In particular, SVM obtains the best results compared to OLS, RR and NN. With a learned representation, LRR+SVM improves the performance of SVM. Besides, GLRR+SVM is capable of boosting the results of LRR+SVM by adding a graph regularization into the original LRR.

For the MISR data set, Fig. 4 and Fig. 5 show the average performances and deviations of seven different models from ten experiments. From these figures, we can observe that the performances of the regression models are better than that of the physical model. In particular, SVM is superior to OLS, RR and NN. GLRR+SVM and LRR+SVM are both better than SVM, because they can learn an effective representation rather than directly using the corrupted spectral values. More importantly, when graph regularization based on the physical retrieval results is incorporated into the LRR model, the performance can be further boosted especially when the number of training samples is less than 70%, which certify the effectiveness of GLRR+SVM.

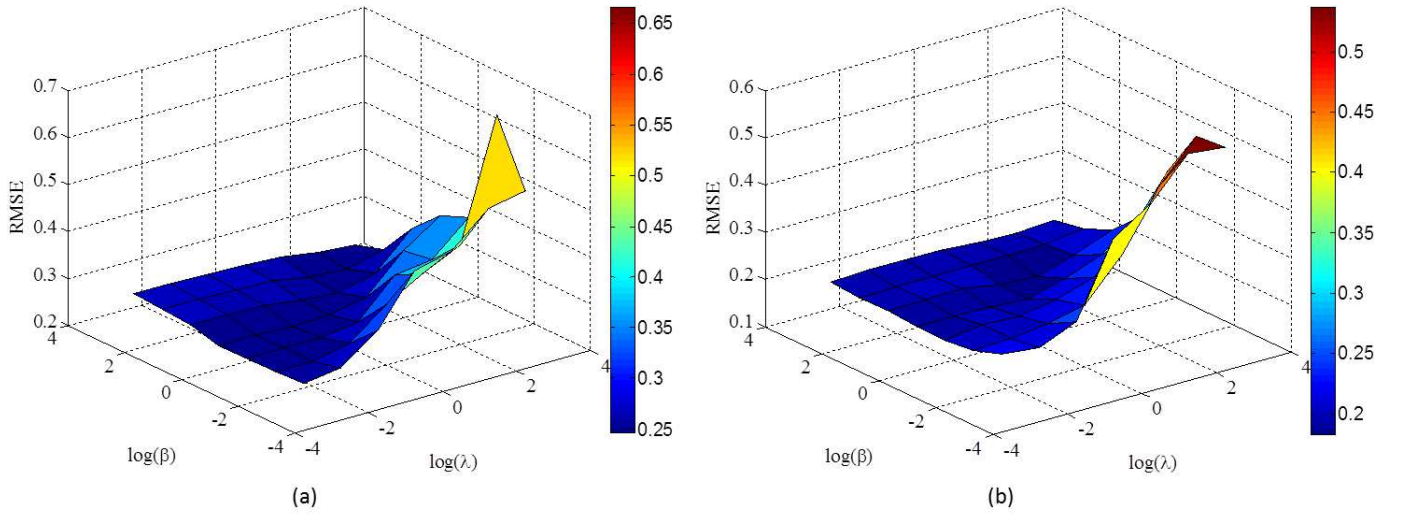


Fig. 6: RMSE of the proposed method versus two different regularization parameters λ and β on (a) MODIS data set. (b) MISR data set.

D. Parameter Analysis

There are two important regularization parameters λ and β in the proposed method. The first one is used to balance the effects of low rank property and noise component, and the second one is utilized to balance the information from empirical model and physical model. Fig. 6 demonstrates the 3-D diagram of RMSE against them on MODIS and MISR data sets. From Fig. 6(a), we can observe that as λ and β increase, the corresponding RMSE firstly increases and then decreases, and achieves the maximal value at $\lambda = 1$ and $\beta = 10^{-1}$. Similar results can be seen from Fig. 6(b), and the maximal value appears when $\lambda = \beta = 10$.

IV. CONCLUSION

This paper proposed a graph regularized low rank representation (LRR) model to learn an effective feature representation for aerosol optical depth (AOD) retrieval. Based on the observation that remote-sensing data often lie in a low rank subspace, LRR was naturally used to uncover the intrinsic data structure from corrupted or noisy observations. Additionally, to preserve the nonlinear structure in remote-sensing data, graph regularization was added to the LRR model. It is well known that the current operating satellites can generate AOD values by physical models. Thus, the proposed graph model was constructed based on such rich information. By conducting experiments on two data sets collected by different instruments

(MODIS and MISR), we compared the proposed method with the physical models and the other state-of-the-art empirical models. The experimental results indicate that the learned representation can improve the retrieval performance.

REFERENCES

- [1] Yoram J Kaufman, Didier Tanré, and Olivier Boucher, “A satellite view of aerosols in the climate system,” *Nature*, vol. 419, no. 6903, pp. 215–223, 2002.
- [2] Lorraine A Remer, YJ Kaufman, D Tanré, S Mattoo, DA Chu, J Vanderlei Martins, R-R Li, C Ichoku, RC Levy, RG Kleidman, et al., “The modis aerosol algorithm, products, and validation,” *Journal of the atmospheric sciences*, vol. 62, no. 4, pp. 947–973, 2005.
- [3] Ralph A Kahn, Barbara J Gaitley, John V Martonchik, David J Diner, Kathleen A Crean, and Brent Holben, “Multiangle imaging spectroradiometer (mISR) global aerosol optical depth validation based on 2 years of coincident aerosol robotic network (aeronet) observations,” *Journal of Geophysical Research: Atmospheres (1984–2012)*, vol. 110, no. D10, 2005.
- [4] YJ Kaufman, D Tanré, L Ao Remer, EF Vermote, A Chu, and BN Holben, “Operational remote sensing of tropospheric aerosol over land from eos moderate resolution imaging spectroradiometer,” *Journal of Geophysical Research: Atmospheres (1984–2012)*, vol. 102, no. D14, pp. 17051–17067, 1997.
- [5] John V Martonchik, David J Diner, Ralph Kahn, Thomas P Ackerman, Michel M Verstraete, Bernard Pinty, Howard R Gordon, et al., “Techniques for the retrieval of aerosol properties over land and ocean using multiangle imaging,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 4, pp. 1212–1227, 1998.
- [6] Robert C Levy, Lorraine A Remer, Shana Mattoo, Eric F Vermote, and Yoram J Kaufman, “Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of moderate resolution imaging spectroradiometer spectral reflectance,” *Journal of Geophysical Research: Atmospheres (1984–2012)*, vol. 112, no. D13, 2007.
- [7] Thi Nhat Thanh Nguyen, Simone Mantovani, and Piero Campalani, “Validation of support vector regression in deriving aerosol optical thickness maps at 1 km² spatial resolution from satellite observations,” in *IEEE International Symposium on Signal Processing and Information Technology*, 2011, pp. 551–556.
- [8] Kosta Ristovski, Slobodan Vucetic, and Zoran Obradovic, “Uncertainty analysis of neural-network-based aerosol retrieval,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 409–414, 2012.
- [9] Slobodan Vucetic, Bo Han, Wen Mi, Zhanqing Li, and Zoran Obradovic, “A data-mining approach for the validation of aerosol retrievals,” *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 1, pp. 113–117, 2008.
- [10] DJ Lary, LA Remer, D MacNeill, B Roscoe, and S Paradise, “Machine learning and bias correction of modis aerosol optical depth,” *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 4, pp. 694–698, 2009.

- [11] Xuefeng Liu, Salah Bourennane, and Caroline Fossati, "Nonwhite noise reduction in hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 3, pp. 368–372, 2012.
- [12] Hongyan Zhang, Liangpei Zhang, and Huanfeng Shen, "A super-resolution reconstruction algorithm for hyperspectral images," *Signal Processing*, vol. 92, no. 9, pp. 2082–2096, 2012.
- [13] Guangcan Liu, Zhouchen Lin, and Yong Yu, "Robust subspace segmentation by low-rank representation," in *Proceedings of the international conference on machine learning*, 2010, pp. 663–670.
- [14] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2013.
- [15] Hongyan Zhang, Wei He, Liangpei Zhang, Huanfeng Shen, and Qiangqiang Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4729–4743, 2014.
- [16] Andrew E Waters, Aswin C Sankaranarayanan, and Richard Baraniuk, "Sparcs: Recovering low-rank and sparse matrices from compressive measurements," in *Advances in neural information processing systems*, 2011, pp. 1089–1097.
- [17] Charles M Bachmann, Thomas L Ainsworth, and Robert A Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 441–454, 2005.
- [18] Miao Zheng, Jiajun Bu, Chun Chen, Can Wang, Lijun Zhang, Guang Qiu, and Deng Cai, "Graph regularized sparse coding for image representation," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1327–1336, 2011.
- [19] Shenghua Gao, Ivor Wai-Hung Tsang, and Liang-Tien Chia, "Laplacian sparse coding, hypergraph laplacian sparse coding, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 92–104, 2013.
- [20] Deng Cai, Xiaofei He, Jiawei Han, and Thomas S Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1548–1560, 2011.
- [21] Arif Albayrak, Jennifer Wei, Maksym Petrenko, Christopher Lynnes, and Robert C Levy, "Global bias adjustment for modis aerosol optical thickness using neural network," *Journal of Applied Remote Sensing*, vol. 7, no. 1, pp. 073514–073514, 2013.
- [22] Guangcan Liu and Shuicheng Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *IEEE International Conference on Computer Vision*. IEEE, 2011, pp. 1615–1622.
- [23] Dalton Lunga, Santasriya Prasad, Melba M Crawford, and Ozan Ersoy, "Manifold-learning-based feature extraction for classification of hyperspectral data: a review of advances in manifold learning," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 55–66, 2014.
- [24] Li Ma, Melba M Crawford, and Jinwen Tian, "Local manifold learning-based-nearest-neighbor for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4099–4109, 2010.
- [25] Wei Li, Saurabh Prasad, James E Fowler, and Lori Mann Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 4, pp. 1185–1198, 2012.
- [26] Mikhail Belkin and Partha Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in neural*

information processing systems, 2001, vol. 14, pp. 585–591.

- [27] Zhouchen Lin, Risheng Liu, and Zhixun Su, “Linearized alternating direction method with adaptive penalty for low-rank representation,” in *Advances in neural information processing systems*, 2011, pp. 612–620.
- [28] Junfeng Yang and Xiaoming Yuan, “Linearized augmented lagrangian and alternating direction methods for nuclear norm minimization,” *Mathematics of Computation*, vol. 82, no. 281, pp. 301–329, 2013.
- [29] Zhouchen Lin, Minming Chen, and Yi Ma, “The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices,” *arXiv preprint arXiv:1009.5055*, 2010.
- [30] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [31] BN Holben, TF Eck, I Slutsker, D Tanre, JP Buis, A Setzer, E Vermote, JA Reagan, YJ Kaufman, T Nakajima, et al., “Aeronetla federated instrument network and data archive for aerosol characterization,” *Remote sensing of environment*, vol. 66, no. 1, pp. 1–16, 1998.
- [32] M Petrenko, C Ichoku, and G Leptoukh, “Multi-sensor aerosol products sampling system (mapss),” *Atmospheric Measurement Techniques*, vol. 5, no. 5, pp. 913–926, 2012.
- [33] Vincent V Salomonson, WL Barnes, Peter W Maymon, Harry E Montgomery, and Harvey Ostrow, “Modis: Advanced facility instrument for studies of the earth as a system,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 27, no. 2, pp. 145–153, 1989.